

FIG. 1. Parallelization Method

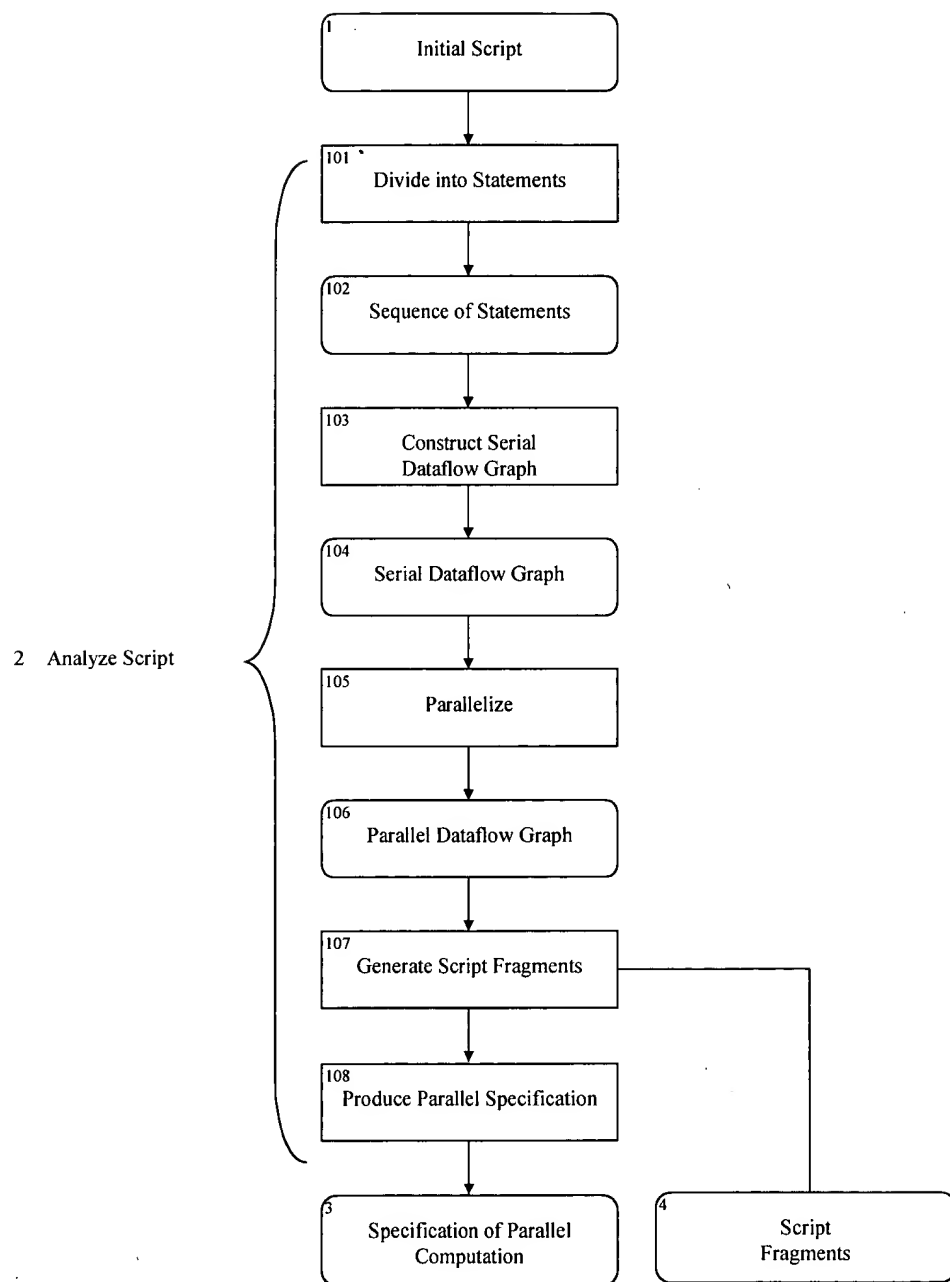


FIG. 2. Method for Analyzing Scripts

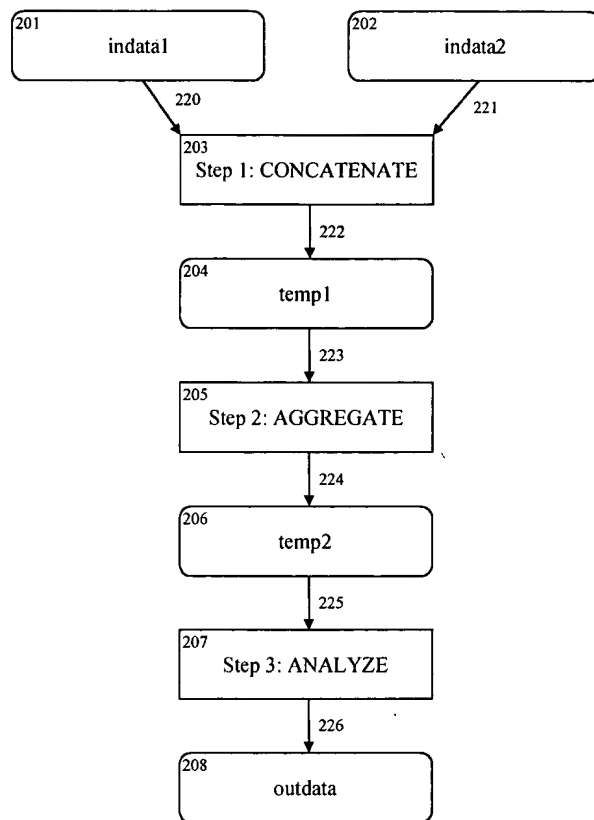


FIG. 3. Serial dataflow graph

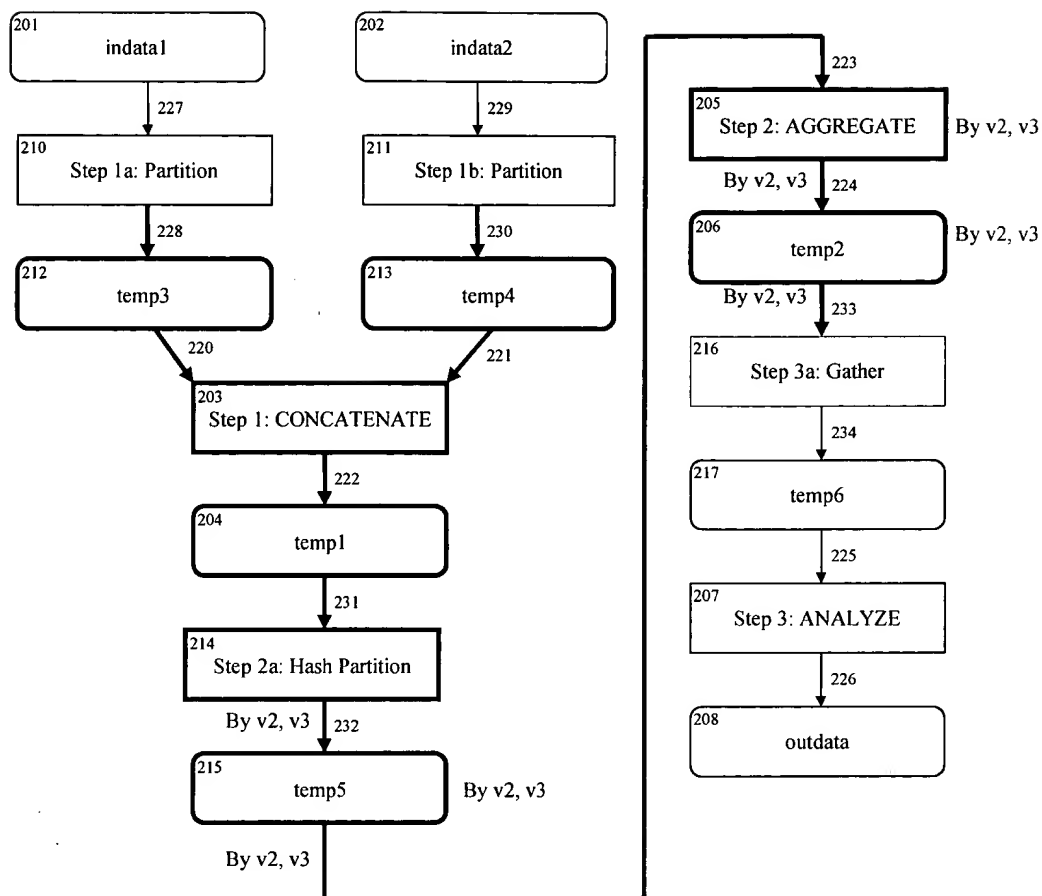


FIG. 4. Parallel Dataflow Graph

	Name	Class	File	Explicit?	
201	indata1	INPUT	input1.dat	Yes	601
202	indata2	INPUT	input2.dat	Yes	
208	outdata	OUTPUT	output.dat	Yes	
204	temp1	TEMP		No	
206	temp2	TEMP		No	
Data Set Table					

	Name	Operation	Parameteres (excepting data sets)	
203	Step1	CONCATENATE		602
205	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	
207	Step3	ANALYZE		
Processing Step Table				

	Step	Data Set	Direction	Role	
220	Step1	indata1	input	in	603
221	Step1	indata2	input	in	
222	Step1	temp1	output	out	
223	Step2	temp1	input	in	
224	Step2	temp2	output	out	
225	Step3	temp2	input	in	
226	Step3	outdata	output	out	
Data Set Access Table					

FIG. 6. Serial Dataset, Processing Step, and Dataset Access Tables

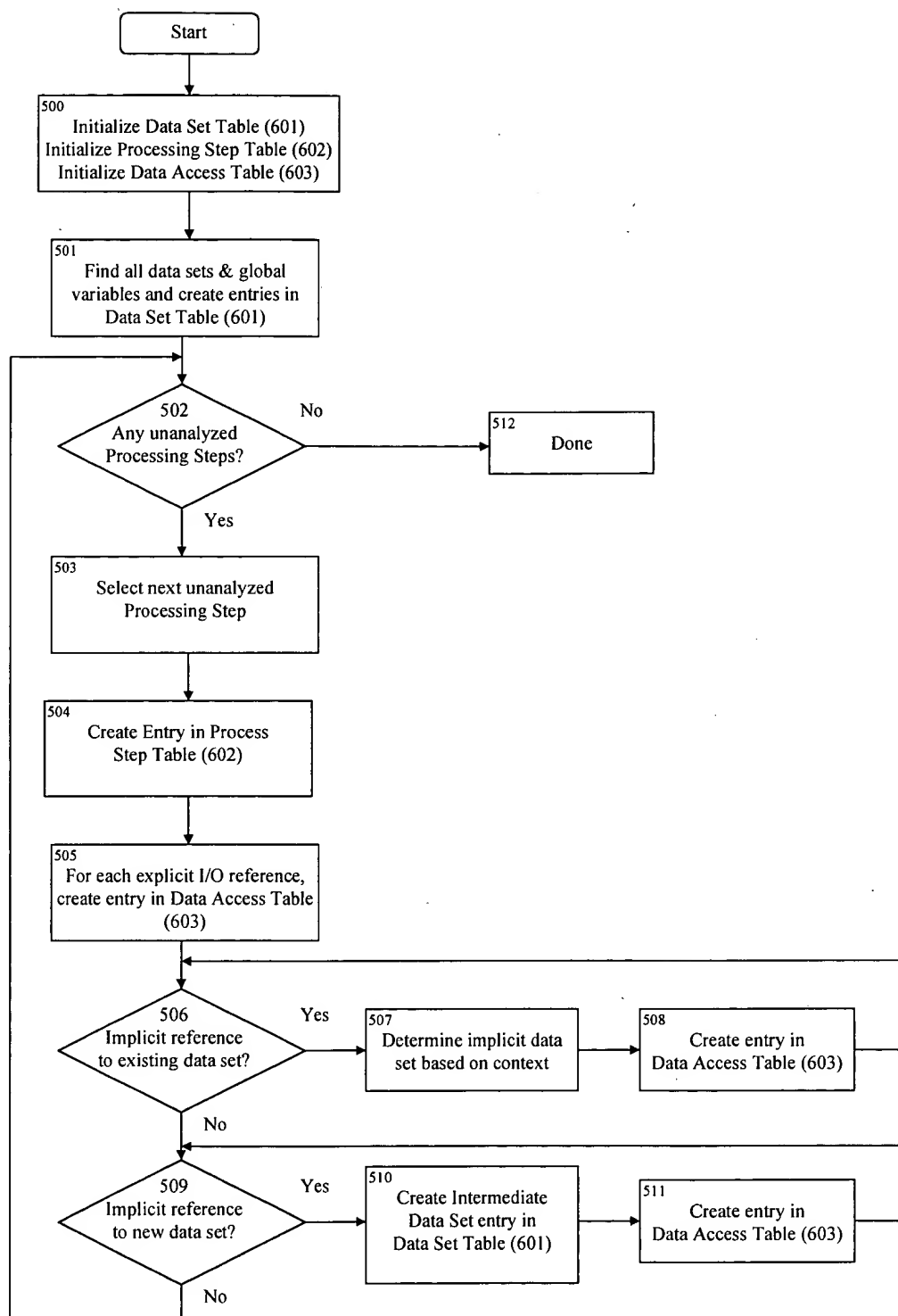


FIG. 7. Method for converting sequences of steps to serial dataflow graph

901	902
903 CONCAT or COPY processing step	Parallelize by simple partitioning
904 Processing step contains BYclause	Parallelize by key-based partitioning
905 INPUT, OUTPUT, or TEMP data set accessing partitioned file	Parallelize by expansion of the partitioned file.
906 AGGREGATE operation	Parallelize by splitting into global and local stages.
907 IN_DB data set accessing partitioned database table.	Parallelize by adding an external "db unload" process.
908 OUT_DB data set accessing partitioned database table	Parallelize by adding an external "db load" process.

FIG. 8. Parallelization repertoire

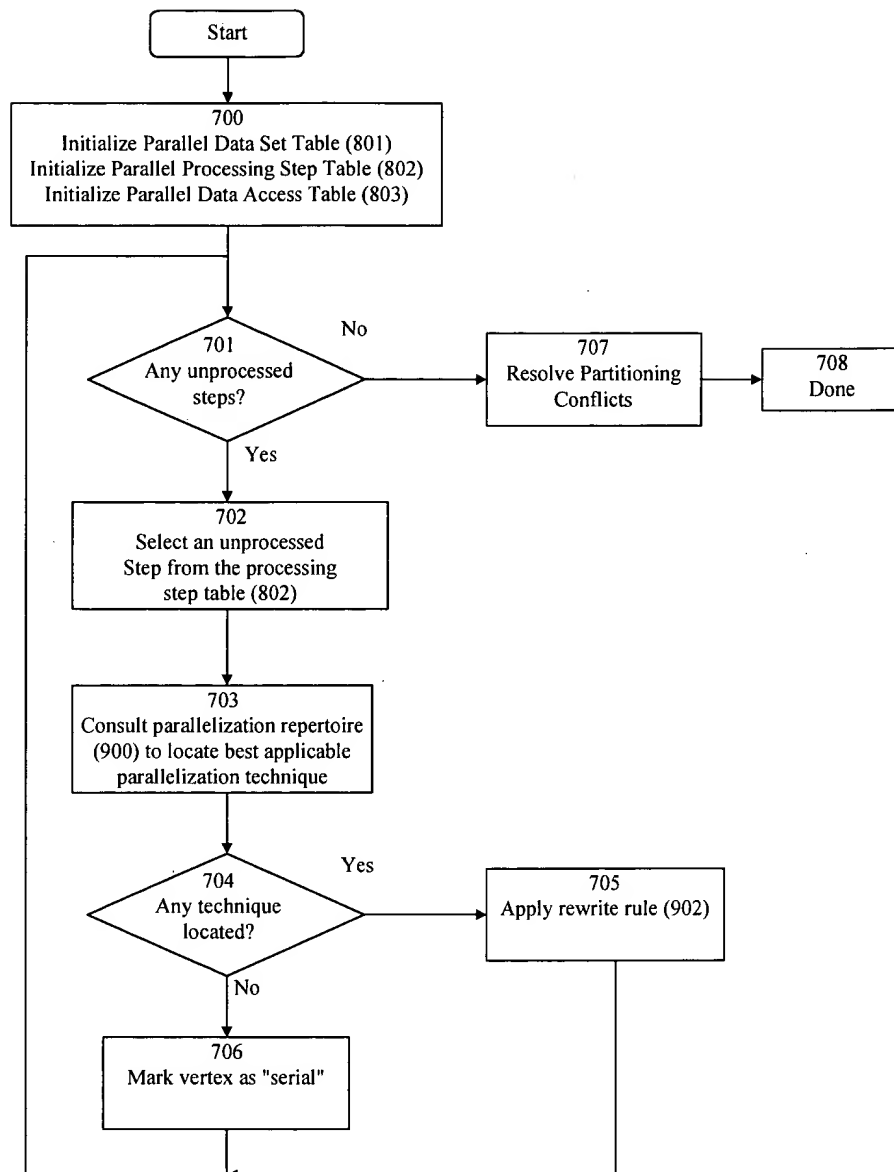


FIG. 9. Method for Parallelizing serial dataflow graph

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	801
206	temp1	TEMP		No		
	temp2	TEMP		No		
	Parallel Data Set Table (initial)					

203	Name	Operation	Parameters (excepting data sets)	Partitioning	
205	Step1	CONCATENATE			
207	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3		802
	Step3	ANALYZE			
	Parallel Processing Step Table (initial)				

220	Step	Data Set	Direction	Role	Partitioning	
221	Step1	indata1	input	in		
222	Step1	indata2	input	in		
223	Step1	temp1	output	out		803
224	Step2	temp1	input	in		
225	Step2	temp2	output	out		
226	Step3	temp2	input	in		
	Step3	outdata	output	out		
	Parallel Data Set Access Table (initial)					

FIG. 10 Initial Parallel Dataset, Processing Step, and Dataset Access Tables

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	801
202	indata2	INPUT	input2.dat	Yes	Serial	
208	outdata	OUTPUT	output.dat	Yes	Serial	
204	temp1	TEMP		No		
206	temp2	TEMP		No		
	Parallel Data Set Table					

203	Name	Operation	Parameters (excepting data sets)	Partitioning	
205	Step1	CONCATENATE		Simple	802
205	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3	
207	Step3	ANALYZE		Serial	
	Parallel Processing Step Table				

220	Step	Data Set	Direction	Role	Partitioning	
221	Step1	indata1	input	in	Simple	803
222	Step1	indata2	input	in	Simple	
222	Step1	temp1	output	out	Simple	
223	Step2	temp1	input	in	By v2, v3	
224	Step2	temp2	output	out	By v2, v3	
225	Step3	temp2	input	in	Serial	
226	Step3	outdata	output	out	Serial	
	Parallel Data Set Access Table					

FIG. 11. Parallel Dataset, Processing Step, and Dataset Access Tables, after parallelizing processing steps.

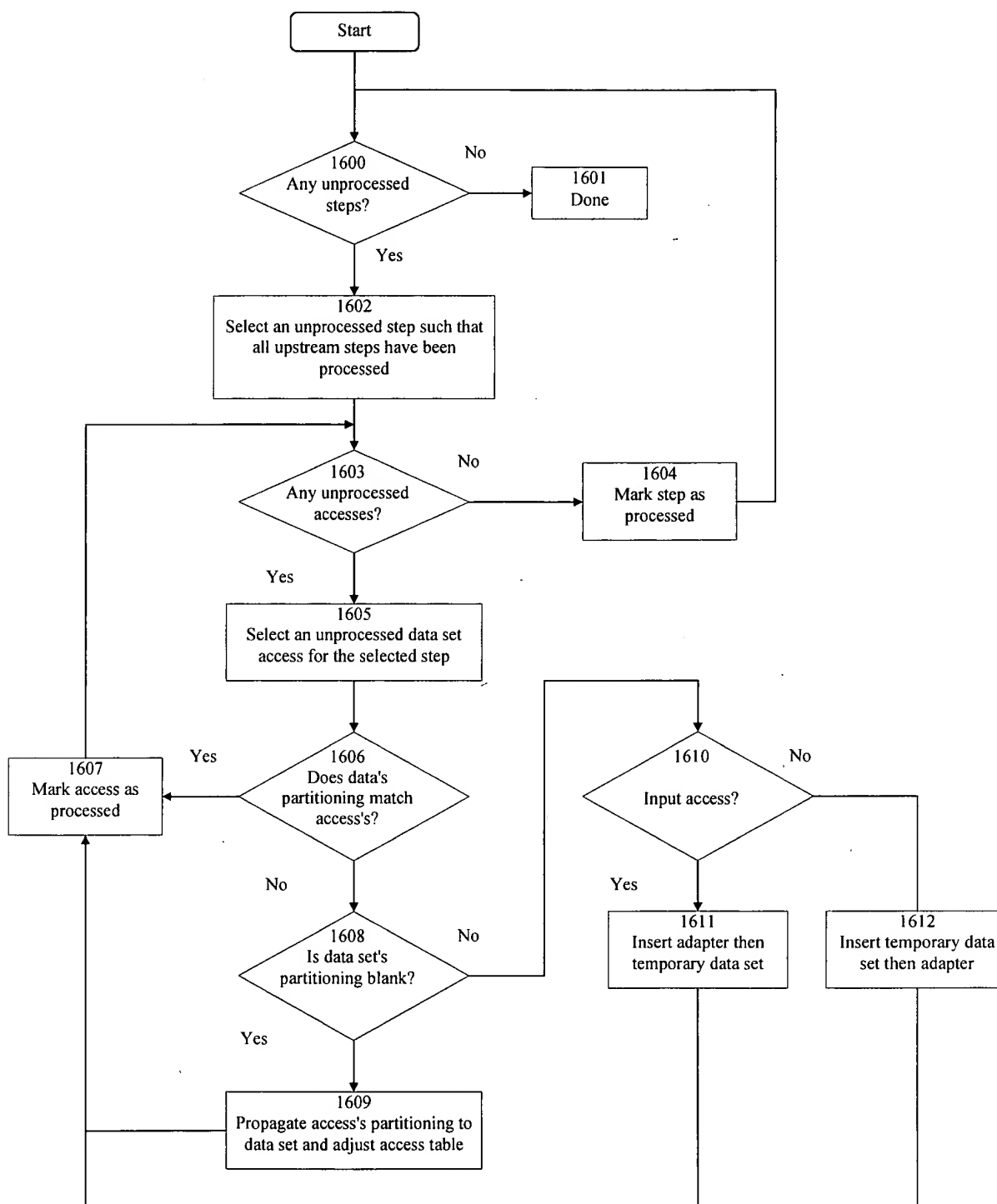


FIG. 12. Method for resolving partitioning conflicts.

	Name	Class	File	Explicit?	Partitioning
201	indata1	INPUT	input1.dat	Yes	Serial
202	indata2	INPUT	input2.dat	Yes	Serial
208	outdata	OUTPUT	output.dat	Yes	Serial
204	temp1	TEMP		No	Simple
206	temp2	TEMP		No	
212	temp3	TEMP		No	Simple
213	temp4	TEMP		No	Simple
Parallel Data Set Table					

	Name	Operation	Parameters (excepting data sets)	Partitioning
210	Step1a	Simple-Partition		Serial
211	Step1b	Simple-Partition		Serial
203	Step1	CONCATENATE		Simple
205	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3
207	Step3	ANALYZE		Serial
Parallel Processing Step Table				

	Step	Data Set	Direction	Role	Partitioning
227	Step1a	indata1	input	in	Serial
228	Step1a	temp3	output	out	Simple
229	Step1b	indata2	input	in	Serial
230	Step1b	temp4	output	out	Simple
220	Step1	temp3	input	in	Simple
221	Step1	temp4	input	in	Simple
222	Step1	temp1	output	out	Simple
223	Step2	temp1	input	in	By v2, v3
224	Step2	temp2	output	out	By v2, v3
225	Step3	temp2	input	in	Serial
226	Step3	outdata	output	out	Serial
Parallel Data Set Access Table					

FIG. 13. Parallel Dataset, Processing Step, and Dataset Access Tables, after resolving Step 1

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	801
206	temp1	TEMP		No	Simple	
212	temp2	TEMP		No	By v2,v3	
213	temp3	TEMP		No	Simple	
215	temp4	TEMP		No	Simple	
	temp5	TEMP		No	By v2,v3	
	Parallel Data Set Table					

210	Name	Operation	Parameters (excepting data sets)	Partitioning	
211	Step1a	Simple-Partition		Serial	
203	Step1b	Simple-Partition		Serial	
214	Step1	CONCATENATE		Simple	
205	Step2a	Hash-Partition	By v2, v3	Simple	
207	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3	
	Step3	ANALYZE		Serial	
	Parallel Processing Step Table				802

227	Step	Data Set	Direction	Role	Partitioning	
228	Step1a	indata1	input	in	Serial	
229	Step1a	temp3	output	out	Simple	
230	Step1b	indata2	input	in	Serial	
220	Step1b	temp4	output	out	Simple	
221	Step1	temp3	input	in	Simple	
222	Step1	temp4	input	in	Simple	
231	Step1	temp1	output	out	Simple	
232	Step2a	temp1	input	in	Simple	
223	Step2a	temp5	output	out	By v2,v3	
224	Step2	temp5	input	in	By v2,v3	
225	Step2	temp2	output	out	By v2,v3	
226	Step3	temp2	input	in	Serial	
	Step3	outdata	output	out	Serial	803
	Parallel Data Set Access Table					

FIG. 14. Parallel Dataset, Processing Step, and Dataset Access Tables, after resolving Step 2

	Name	Class	File	Explicit?	Partitioning
201	indata1	INPUT	input1.dat	Yes	Serial
202	indata2	INPUT	input2.dat	Yes	Serial
208	outdata	OUTPUT	output.dat	Yes	Serial
204	temp1	TEMP		No	Simple
206	temp2	TEMP		No	By v2,v3
212	temp3	TEMP		No	Simple
213	temp4	TEMP		No	Simple
215	temp5	TEMP		No	By v2,v3
217	temp6	TEMP		No	Serial
Parallel Data Set Table					

	Name	Operation	Parameters (excepting data sets)	Partitioning
210	Step1a	Simple-Partition		Serial
211	Step1b	Simple-Partition		Serial
203	Step1	CONCATENATE		Simple
214	Step2a	Hash-Partition	By v2, v3	Simple
205	Step2	AGGREGATE	o1 = min v1, o2 = max v1, o3 = avg v1 BY v2, v3	By v2,v3
216	Step3a	Gather		Serial
207	Step3	ANALYZE		Serial
Parallel Processing Step Table				

	Step	Data Set	Direction	Role	Partitioning
227	Step1a	indata1	input	in	Serial
228	Step1a	temp3	output	out	Simple
229	Step1b	indata2	input	in	Serial
230	Step1b	temp4	output	out	Simple
220	Step1	temp3	input	in	Simple
221	Step1	temp4	input	in	Simple
222	Step1	temp1	output	out	Simple
231	Step2a	temp1	input	in	Simple
232	Step2a	temp5	output	out	By v2,v3
223	Step2	temp5	input	in	By v2,v3
224	Step2	temp2	output	out	By v2,v3
233	Step3a	temp2	input	in	By v2,v3
234	Step3a	temp6	output	out	Serial
225	Step3	temp6	input	in	Serial
226	Step3	outdata	output	out	Serial
Parallel Data Set Access Table					

FIG. 15. Parallel Dataset, Processing Step, and Dataset Access Tables, after resolving Step 3

Figure	Chemical Structure	Yield (%)	mp (°C)	lit. mp (°C)	lit. yield (%)
1		70	100-101	100-101	70
2		70	100-101	100-101	70
3		70	100-101	100-101	70
4		70	100-101	100-101	70
5		70	100-101	100-101	70
6		70	100-101	100-101	70
7		70	100-101	100-101	70
8		70	100-101	100-101	70
9		70	100-101	100-101	70
10		70	100-101	100-101	70
11		70	100-101	100-101	70
12		70	100-101	100-101	70
13		70	100-101	100-101	70
14		70	100-101	100-101	70
15		70	100-101	100-101	70
16		70	100-101	100-101	70
17		70	100-101	100-101	70
18		70	100-101	100-101	70
19		70	100-101	100-101	70
20		70	100-101	100-101	70
21		70	100-101	100-101	70
22		70	100-101	100-101	70
23		70	100-101	100-101	70
24		70	100-101	100-101	70
25		70	100-101	100-101	70
26		70	100-101	100-101	70
27		70	100-101	100-101	70
28		70	100-101	100-101	70
29		70	100-101	100-101	70
30		70	100-101	100-101	70
31		70	100-101	100-101	70
32		70	100-101	100-101	70
33		70	100-101	100-101	70
34		70	100-101	100-101	70
35		70	100-101	100-101	70
36		70	100-101	100-101	70
37		70	100-101	100-101	70
38		70	100-101	100-101	70
39		70	100-101	100-101	70
40		70	100-101	100-101	70
41		70	100-101	100-101	70
42		70	100-101	100-101	70

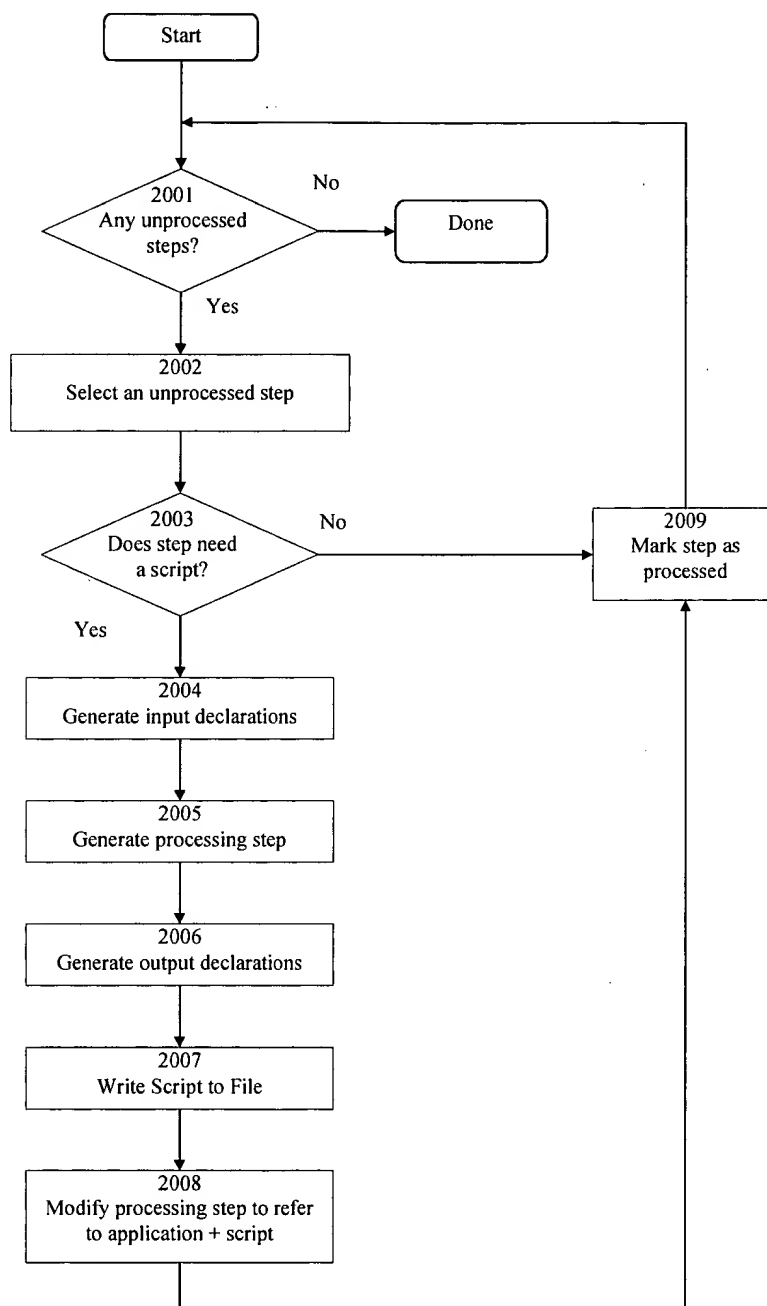


FIG. 16. Method for generating script fragments

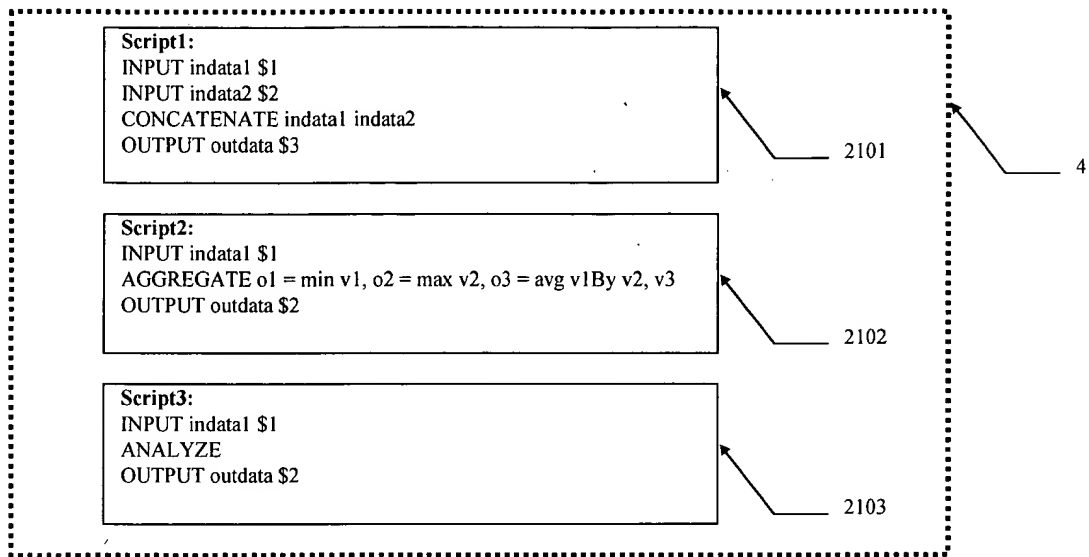
[illegible]

FIG. 17. Script Fragments File

	Name	Class	File	Explicit?	Partitioning	
201	indata1	INPUT	input1.dat	Yes	Serial	801
202	indata2	INPUT	input2.dat	Yes	Serial	
208	outdata	OUTPUT	output.dat	Yes	Serial	
204	temp1	TEMP		No	Simple	
206	temp2	TEMP		No	By v2,v3	
212	temp3	TEMP		No	Simple	
213	temp4	TEMP		No	Simple	
215	temp5	TEMP		No	By v2,v3	
217	temp6	TEMP		No	Serial	
Parallel Data Set Table						

	Name	Operation	Parameters (excepting data sets)	Partitioning	
210	Step1a	Simple-Partition		Serial	802
211	Step1b	Simple-Partition		Serial	
203	Step1	Run analyze	script1	Simple	
214	Step2a	Hash-Partition	By v2, v3	Simple	
205	Step2	Run analyze	script2	By v2,v3	
216	Step3a	Gather		Serial	
207	Step3	Run analyze	script3	Serial	
Parallel Processing Step Table					

	Step	Data Set	Direction	Role	Partitioning	
227	Step1a	indata1	input	in	Serial	803
228	Step1a	temp3	output	out	Simple	
229	Step1b	indata2	input	in	Serial	
230	Step1b	temp4	output	out	Simple	
220	Step1	temp3	input	in	Simple	
221	Step1	temp4	input	in	Simple	
222	Step1	temp1	output	out	Simple	
231	Step2a	temp1	input	in	Simple	
232	Step2a	temp5	output	out	By v2,v3	
223	Step2	temp5	input	in	By v2,v3	
224	Step2	temp2	output	out	By v2,v3	
233	Step3a	temp2	input	in	By v2,v3	
234	Step3a	temp6	output	out	Serial	
225	Step3	temp6	input	in	Serial	
226	Step3	outdata	output	out	Serial	
Parallel Data Set Access Table						

FIG. 18. Parallel Dataset, Processing Step, and Dataset Access Tables, after generation of script fragments

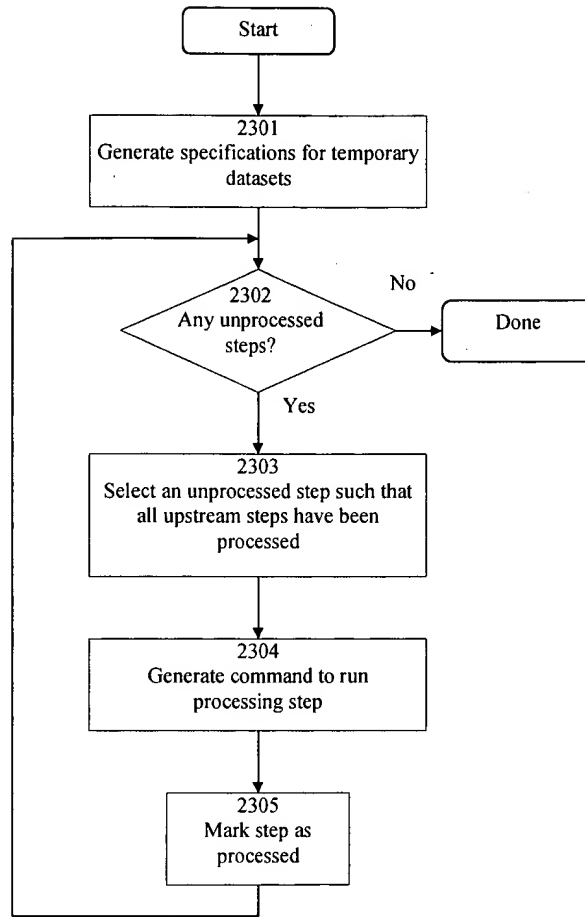


FIG. 19. Method for generating parallel specification

201	Name	Class	File	Explicit?	Partitioning	
202	indata1	INPUT	input1.dat	Yes	Serial	
208	indata2	INPUT	input2.dat	Yes	Serial	
204	outdata	OUTPUT	output.dat	Yes	Serial	801
206	temp1	TEMP	temp1.1;temp1.2	No	Simple	
212	temp2	TEMP	temp2.1;temp2.2	No	By v2,v3	
213	temp3	TEMP	temp3.1;temp3.2	No	Simple	
215	temp4	TEMP	temp4.1;temp4.2	No	Simple	
217	temp5	TEMP	temp5.1;temp5.2	No	By v2,v3	
	temp6	TEMP	temp6	No	Serial	
	Parallel Data Set Table					

210	Name	Operation	Parameters (excepting data sets)	Partitioning	
211	Step1a	Simple-Partition		Serial	
203	Step1b	Simple-Partition		Serial	
214	Step1	run analyze	script1	Simple	802
205	Step2a	Hash-Partition	By v2, v3	Simple	
216	Step2	run analyze	script2		
207	Step3a	Gather		Serial	
	Step3	run analyze	script3	Serial	
	Parallel Processing Step Table				

227	Step	Data Set	Direction	Role	Partitioning	
228	Step1a	indata1	input	in	Serial	
229	Step1a	temp3	output	out	Simple	
230	Step1b	indata2	input	in	Serial	
220	Step1b	temp4	output	out	Simple	
221	Step1	temp3	input	in	Simple	
222	Step1	temp4	input	in	Simple	
231	Step1	temp1	output	out	Simple	
232	Step2a	temp1	input	in	Simple	
223	Step2a	temp5	output	out	By v2,v3	
224	Step2	temp5	input	in	By v2,v3	
233	Step2	temp2	output	out	By v2,v3	
234	Step3a	temp2	input	in	By v2,v3	
225	Step3a	temp6	output	out	Serial	
226	Step3	temp6	input	in	Serial	
	Step3	outdata	output	out	Serial	
	Parallel Data Set Access Table					803

FIG. 20. Parallel Dataset, Processing Step, and Dataset Access Tables, after defining temporary datasets

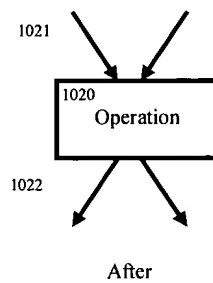
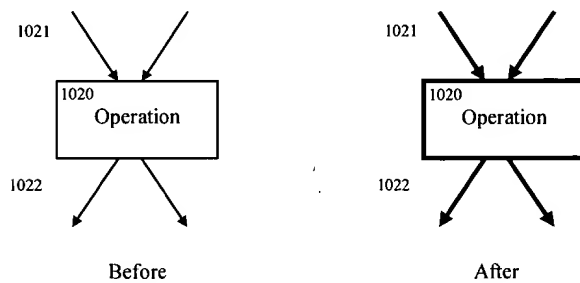
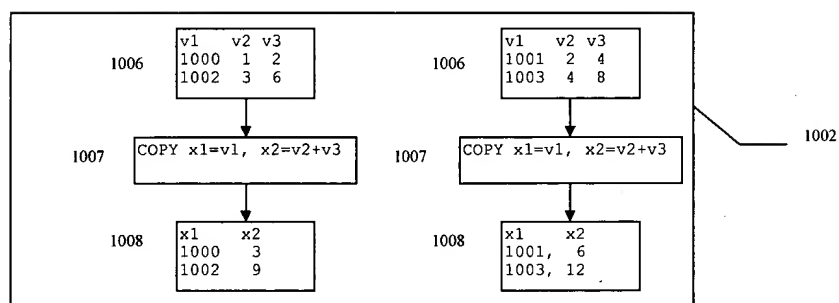
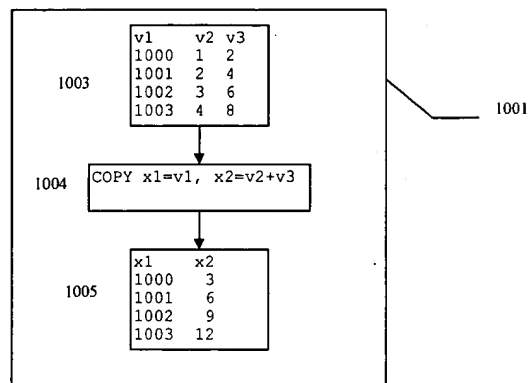
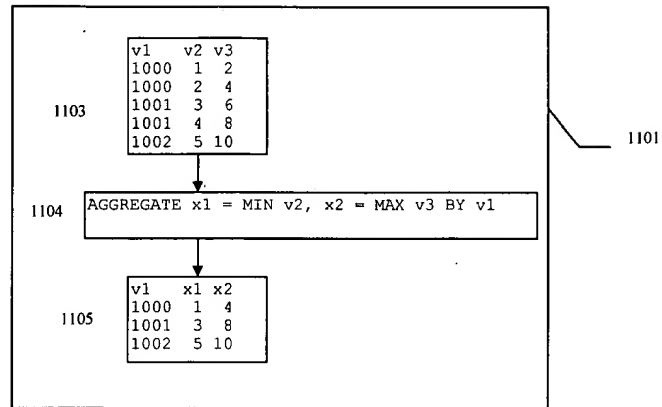
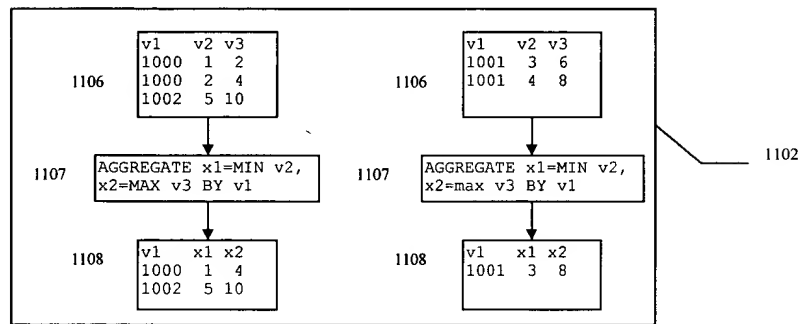


FIG. 22. Simple Partitioning



Serial Version



Parallel Version

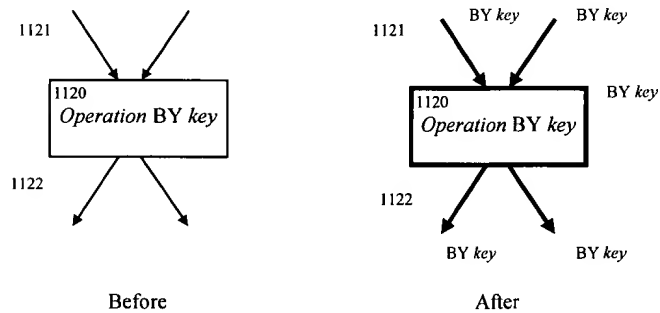
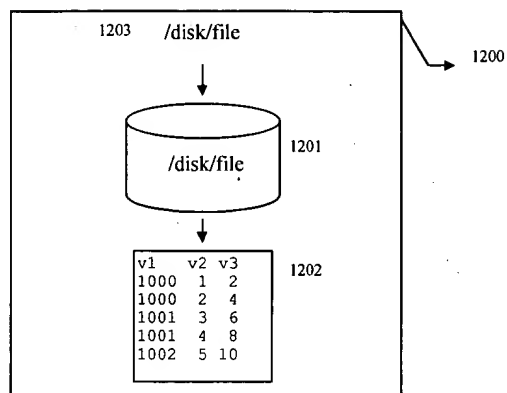
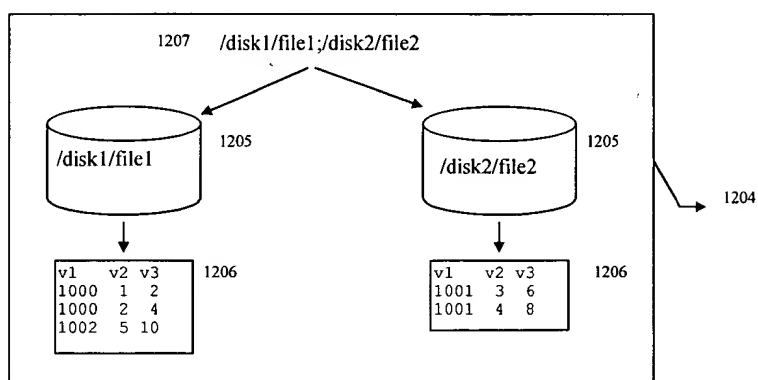


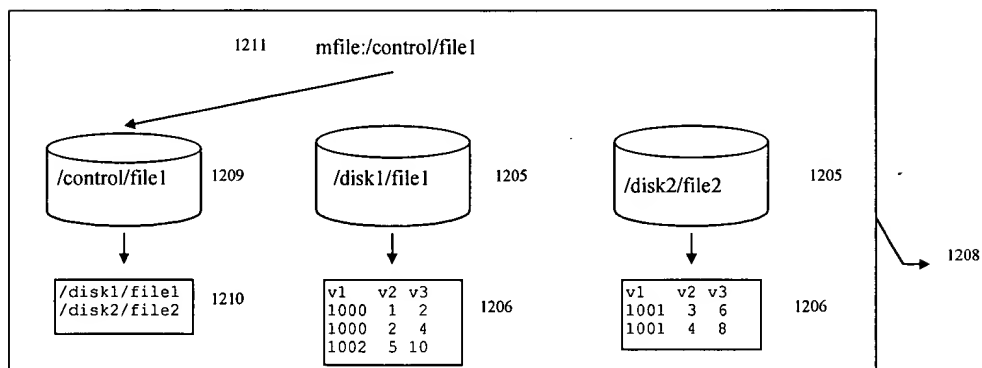
FIG. 23. Key-based Partitioning



Serial Dataset



Partitioned Dataset



Partitioned Dataset with Parallel Virtual File System

FIG. 24. Access to Partitioned Datasets

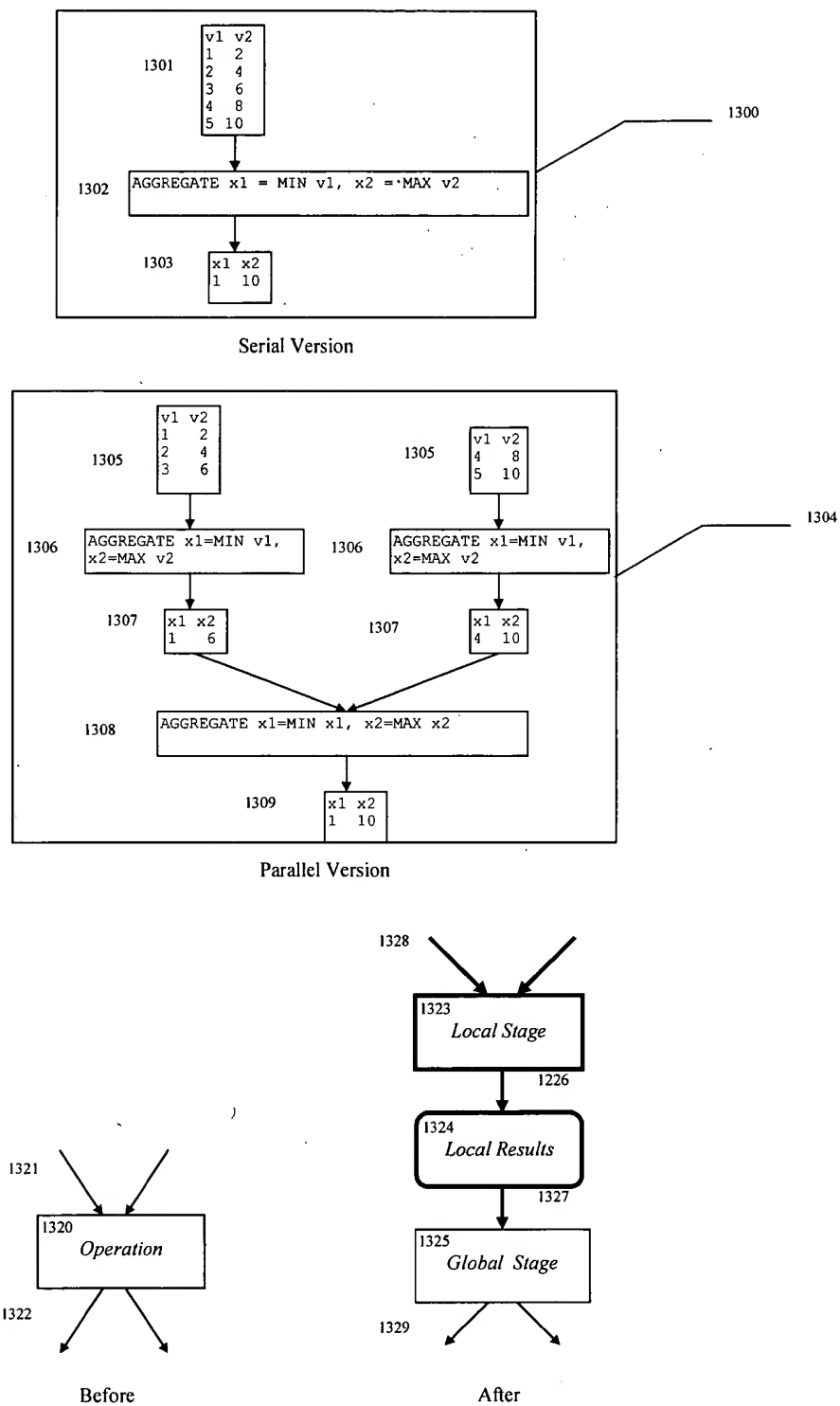


FIG. 25. Local-Global Division Parallelization

602770-54862250

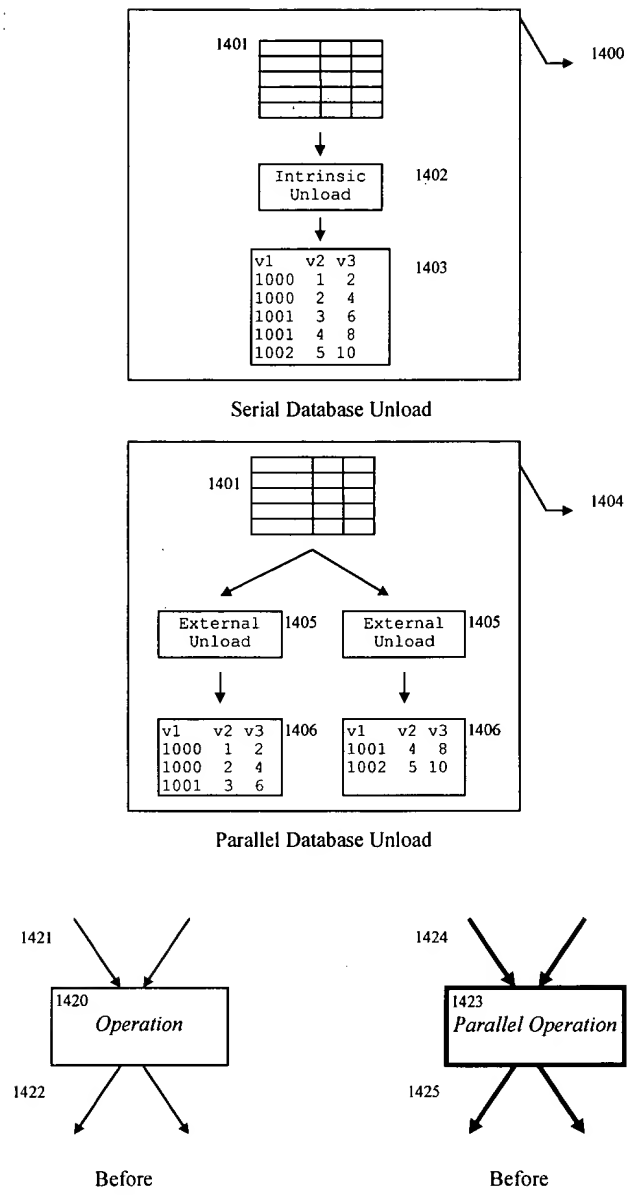


FIG. 26. External Parallelism

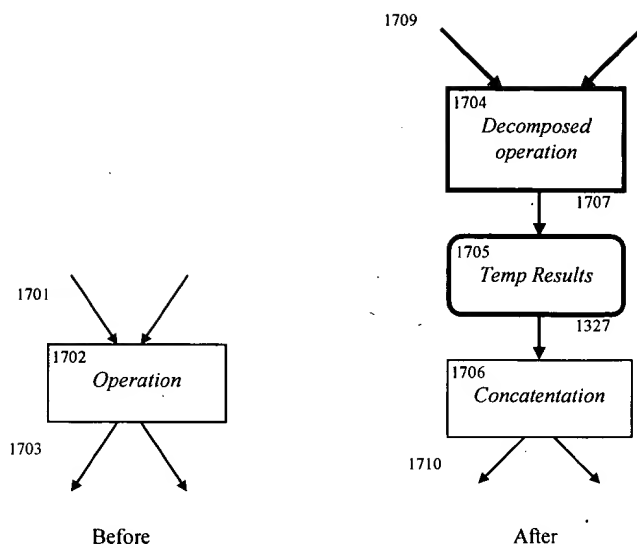


FIG. 27. Statement Decomposition

66EFD: 678626B

```
graph TD
    subgraph Before
        I1501_1[1501] --> B1505[1505 MEAN Procedure]
        B1505 --> O1502_1[1502]
    end
    subgraph After
        I1501_2[1501] --> B1504[1504 Partitioned Dataset]
        B1504 --> B1505_2[1505 MEAN Procedure]
        B1505_2 --> B1506[1506 Local Datasets]
        B1506 --> B1507[1507 External Merge Application]
        B1507 --> O1502_2[1502]
    end
```

```

graph TD
    subgraph Before
        1601a[1601] --> 1605a[1605  
FREQ Procedure]
        1605a --> 1602a[1602]
    end
    subgraph After
        1601b[1601] --> 1604b[1604  
Partitioned Dataset]
        1604b --> 1605b[1605  
FREQ Procedure]
        1605b --> 1606b[1606  
Local Datasets]
        1606b --> 1607b[1607  
FREQ Procedure]
        1607b --> 1602b[1602]
    end

```

Before

After

FIG. 29. Local-Global Division for SAS FREQ process

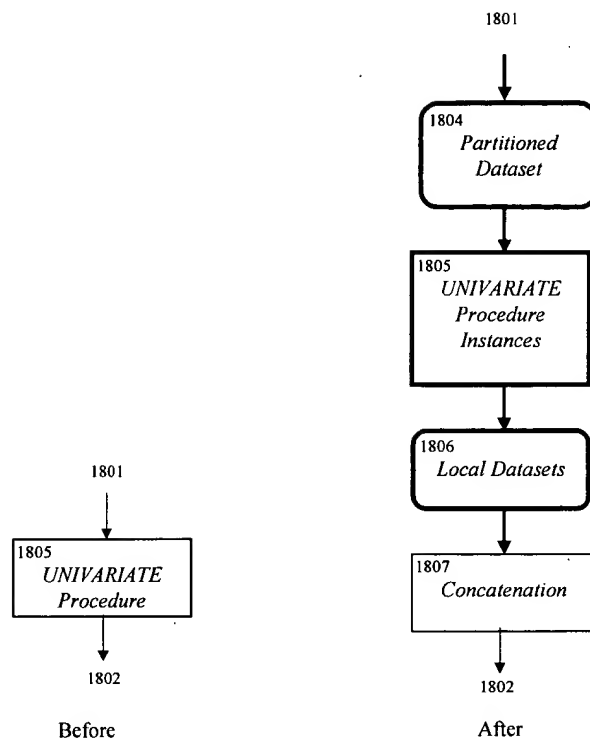


FIG. 30. Statement Decomposition for SAS UNIVARIATE process